

Présentation

Sylvie Mellet

**Édition électronique**

URL : <http://journals.openedition.org/corpus/2148>

DOI : 10.4000/corpus.2148

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Édition imprimée

Date de publication : 1 novembre 2011

Pagination : 7-11

ISSN : 1638-9808

Référence électronique

Sylvie Mellet, « Présentation », *Corpus* [En ligne], 10 | 2011, mis en ligne le 18 juin 2012, consulté le 21 septembre 2020. URL : <http://journals.openedition.org/corpus/2148> ; DOI : <https://doi.org/10.4000/corpus.2148>

© Tous droits réservés

Présentation

Sylvie MELLET

BCL, CNRS, Université Nice Sophia Antipolis

Pour sa dixième livraison, la revue **CORPUS** offre à ses lecteurs, après neuf numéros strictement thématiques, un volume de *varia*. Le comité de rédaction a souhaité en effet laisser le champ libre aux soumissions spontanées, de plus en plus nombreuses, et par là même donner un aperçu diversifié des recherches actuelles dans le champ couvert par la revue.

En réalité, en dépit de cette ouverture, cette livraison ne saurait prétendre à l'exhaustivité : nous n'avons reçu, par exemple, aucune proposition d'article sur les problèmes afférents à l'utilisation du web comme corpus¹ ; nous n'avons pu retenir aucune des propositions sur les questions méthodologiques de constitution, annotation et exploitation de corpus aligné, ou de corpus interlangue et interculturel.

En revanche, ce numéro témoigne d'une permanence, voire d'une densification, de la réflexion des linguistes pour s'approprier l'objet 'corpus' et en faire un véritable observatoire spécifique à la discipline. Cette appropriation passe par une analyse des besoins en ressources (diversification des données, enrichissement des corpus bruts, amélioration des outils d'exploitation), par une confrontation avec l'usage des autres disciplines, enfin par un retour critique sur l'apport des corpus lorsqu'ils permettent de réévaluer certaines assertions de la linguistique théorique.

¹ Peut-être parce que les contributeurs potentiels avaient préféré travailler pour la *Revue de linguistique française et d'analyse du discours*, dont le tome 2.1 du printemps 2011 est consacré à « L'internet, corpus sauvage. Nouvelles ressources, nouveaux problèmes ? ».

Le numéro est donc organisé, *grosso modo*, selon ces trois axes.

Dans une première partie, le propos sera assez technique car consacré à la présentation de procédures d'annotation nouvelles, soit parce que le type des corpus eux-mêmes est assez récent dans le champ de la linguistique de corpus, soit parce que les informations ciblées par l'analyste sont particulièrement complexes. Ainsi, le premier article, d'Annelies BRAFFORT, Emilie CHETELAT-PELE et Jérémie SEGOUAT, analyse les principales étapes de l'annotation d'un corpus de langue des signes et tente de situer les biais possibles des méthodes d'annotation et d'analyse d'un tel corpus. Tsuyoshi KIDA, pour sa part, décrit le modèle d'annotation qu'il a mis au point pour coder l'activité kinésique qui accompagne toute production discursive orale, notamment en interaction. Enfin, dans un registre plus classique, celui d'un corpus de textes écrits, Frédéric LANDRAGIN, qui travaille sur l'annotation des chaînes co-référentielles, expose les problèmes rencontrés et les solutions adoptées face à la nature sémantico-pragmatique des phénomènes étudiés, à leur complexité et à la présence fréquente d'ambiguïtés ou de sous-déterminations.

La deuxième partie commence par présenter de nouveaux corpus et tend surtout à interroger le rapport du linguiste à la matérialité discursive de ses corpus. Lorsque les corpus sont des corpus multi-usages, explorés par d'autres disciplines telles que la sociologie, la politologie, voire la médecine, quelle est l'attitude épistémologique particulière que peut avancer le linguiste, notamment dans le cadre de collaborations interdisciplinaires ? Plus modestement, quelles sont les propriétés attendues et nécessaires d'un corpus sociolinguistique pour la recherche linguistique ? Telles sont quelques-unes des questions auxquelles cherchent à répondre Sonia BRANCA-ROSOFF, Serge FLEURY, Florence LEFEUVRE et Matthew PIRES en présentant le CFPP2000 (Corpus du français parlé parisien des années 2000) ou Olivier BAUDE et Céline DUGUA en confrontant deux corpus d'enquêtes sociolinguistiques réalisés à quarante ans d'intervalle dans la région d'Orléans. Sandra NOSSIK généralise cette réflexion épistémologique à propos

Présentation

d'un corpus de « récits de vie », ce qui la conduit à prôner la complémentarité d'une approche discursive fondée sur la matérialité des discours et d'une approche interactionnelle plus encline à tenir compte des facteurs d'analyse externes, notamment sociologiques. Enfin, deux articles illustrent à travers une étude concrète et précise la façon dont les linguistes peuvent rendre compte d'un corpus qui n'a pas, au départ, été constitué à des fins linguistiques, ni même pour constituer un réservoir de données langagières. A travers une approche outillée d'un corpus de consultations médicales, Ludovic TANGUY, Cécile FABRE, Lydia-Mai HO-DAC et Josette REBEYROLLE montrent comment, en tant que linguistes familiers du traitement outillé des corpus, ils ont abordé ce matériau particulier et quels éléments de caractérisation spécifiques les linguistes sont en mesure d'apporter en réponse à une demande émanant de la sphère médicale. Mathieu BRUGIDOU, pour sa part, analyse les documents issus du Grenelle de l'environnement en tant que dispositif d'écriture et d'énonciation collective.

La troisième partie rassemble des études linguistiques sur corpus. Celle de Lydia-Mai HO-DAC et Anne KÜPPERS a pour objectif de caractériser contrastivement, au regard de l'expression de la subjectivité, un corpus de presse composé de presse écrite traditionnelle et de médias participatifs en ligne. Les trois autres études confrontent des assertions classiquement admises à la réalité des faits telle qu'elle peut être appréhendée, soit à travers l'étude d'un corpus construit pour documenter prototypiquement le problème linguistique sous étude, soit à travers l'analyse des données fournies par un corpus de référence. Le premier cas de figure est illustré par deux articles : celui d'Emeline LECUIT, Denis MAUREL, Duško VITAS soumet l'assertion « les noms propres ne se traduisent pas » à la pratique observable dans un ensemble de traductions en dix langues différentes du roman de Jules Verne *Le tour du monde en 80 jours*, particulièrement riche en noms propres ; les résultats linguistiques sont accompagnés d'une évaluation des atouts et des limites du corpus. L'article de Delphine BEAUSEROY, Evelyne JACQUEY et Marie-Laurence KNITTEL confronte l'hypothèse de Grimshaw (1990) sur les noms déverbaux,

formulée dans le cadre d'une linguistique introspective, aux résultats de tests appliqués aux données fournies par 3 années de livraison d'un quotidien régional ; là encore les résultats sont évalués, et si le corpus apporte des rectificatifs à l'hypothèse initiale, les auteurs reconnaissent aussi volontiers « que l'étude sur corpus soulève certaines difficultés qui nécessitent de la compléter soit en aménageant les exemples, soit en retournant vers des exemples construits ».

Le deuxième cas de figure est mis en pratique dans l'article d'Arian SHAHROKHY-PREHN et Silke HÖCHE, consacré aux « Light Verb Constructions » de l'anglais. Ces constructions sont réputées familières, appréciation qu'il convient de nuancer après examen des données du corpus de référence COCA (Corpus of Contemporary American English).

Ce tour d'horizon des outils et pratiques de la linguistique de corpus prend fin avec deux contributions un peu particulières : on lira tout d'abord la version abrégée d'un article de Bénédicte PINCEMIN (paru en espagnol dans sa version longue) et centré ici sur les nombreuses connivences entre la théorie de la sémantique interprétative et les principes fondateurs de l'approche textométrique. Ainsi se trouve illustré, par un exemple disciplinaire précis, le lien étroit qui peut s'établir, dans une pratique empirique profondément réflexive, entre des méthodes outillées de traitement de corpus, et une théorie linguistique qui, en posant le contexte comme condition des effets de sens, revendique le corpus (textuel) comme objet empirique de la linguistique. Pour finir, la contribution de Frédérique MELANIE-BECQUET et Catherine FUCHS inaugure une nouvelle rubrique de la revue, dont on espère qu'elle pourra se perpétuer, intitulée « Ressources » : il s'agit pour les auteurs de présenter aux lecteurs un corpus nouveau, ainsi que les outils d'exploration spécifiques qui lui ont été associés. Le tout doit être accessible en ligne et, par conséquent, offrir de nouvelles ressources à la communauté. En l'occurrence, il s'agit ici d'une présentation de la base de données construite au LATTICE sur l'expression de la comparaison dans les langues romanes.

Présentation

Le numéro se clôt sur la traditionnelle rubrique des comptes rendus, dont nous remercions sincèrement les rédacteurs.